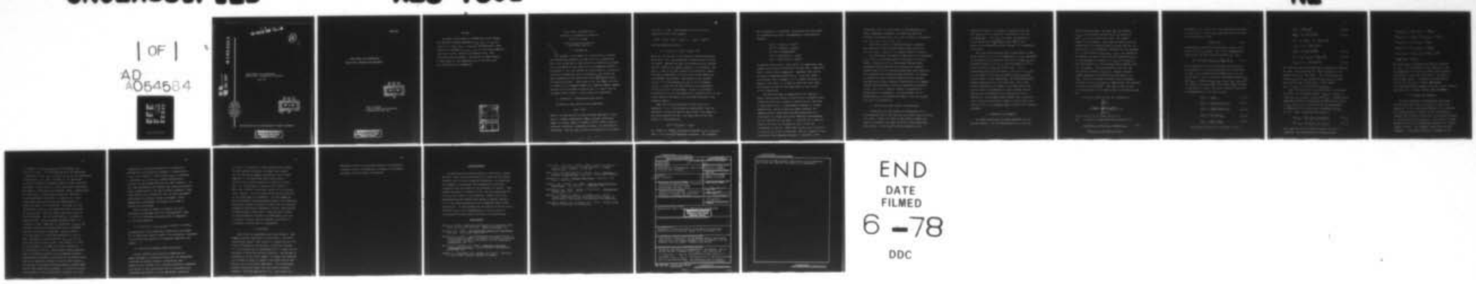
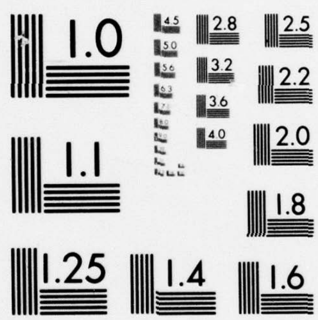
AD-A054 584 UNION COLL SCHENECTADY N Y INST OF ADMINISTRATION AN--ETC F/G 12/1
TIME SERIES IN M-DIMENSIONS: DEFINITION, PROBLEMS AND PROSPECTS--ETC(U)
JAN 78 L A AROIAN N00014-77-C-0438
UNCLASSIFIED AES-7801 NL

1 OF 1
AD
A054584



END
DATE
FILMED
6-78
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

TIME SERIES IN M-DIMENSIONS
DEFINITION, PROBLEMS AND PROSPECTS

AES-7801

ADMINISTRATIVE AND ENGINEEERING SYSTEMS MONOGRAPH

AES-7801

TIME SERIES IN M-DIMENSIONS
DEFINITION, PROBLEMS AND PROSPECTS

Leo A. Aroian
Union College and University
Schenectady, New York

Abstract

We define time series in m dimensions **x**, as follows: the observed variable z depends on x_{1i}, $x_{2i}, \ldots, x_{mi}$ and $t_i \le t$ or $z = f(x_{1i}, x_{2i}, \ldots, x_{mi}, t_i)$, and similarly n time series in m variables $z_1, z_2, \ldots z_n$, $z \equiv f(x)$ where $\mathbf{z}=\mathbf{f}(\mathbf{x},\mathbf{t})$ and x are vectors. This is the discrete case. The continuous case is similar. Distinction is made between m time series in zero dimension, all on the line, and one time series in m dimensions.

Time Series in M-Dimensions

Definition, Problems and Prospects

by Leo A. Aroian

Union College and University
Schenectady, New York

1. Introduction

The purpose of this paper is to define and to consider the problems and prospects of time series in m dimensions. First, the definition of time series in m dimensions, secondly its relationship to time series in zero dimensions, whether stationary or not, and finally the relationship to correlation theory in m dimensions will be considered. The author has been influenced by the basic work of Box and Jenkins (1970) in prediction theory. Their work in the time domain and that of others in the frequency domain has forged an almost complete solution of the problems of time series at a point. We are extending their results to time series in m dimensions.

2. Definition - M-Dimensional Time Series

We define a time series in zero dimensions

$$z(t) = f(t)$$

where t is time and z(t) is some variable dependent on time such as voltage, temperature, wages, economic indicators, number of births, number of deaths, prices, crimes, social, business, industrial, biological, medical, and environmental conditions. Now z(t) may be given continuously or at discrete intervals of time. Additionally there may be m zero dimensional time series namely:

$$z_1(t) = f_1(t),\ z_2(t) = f_2(t),\ \ldots.\ z_m(t) = f_m(t),$$

and interrelations such as

$$z_i = f_i(z_1, z_2, \ldots, z_j, \ldots, z_m; t),\ i \neq j,$$

where the z's are any of the specific variables mentioned previously. The time series may be known exactly at least in theory. Usually they must be estimated and are defined only for $-\infty < t \leq t_o$, or $t_1 < t \leq t_o$, where t_o is the present time and t_1 a time in the past chosen for a particular reason. They may be stationary or if not may be transformed into a stationary time series. Time series of one variable $\{z_1, t\}$, or of two variables $\{z_1, z_2, t\}$, or m variables $\{z_1, z_2, \ldots, z_m, t\}$ are all time series in zero dimensions even in the case of m variables. We may be interested in prediction, in the time domain, i.e. prediction of z in the future based on its past; or in the power spectrum in the frequency domain.

What then is the definition of time series in m dimensions or of n time series in m dimensions? Does the definition include the special case of m=0? The answer to the second question is yes. Our definition of one time series in m dimensions is:

$$z(t) = f(x_1, x_2, \ldots., x_m; t)$$

for $-\infty < t < \infty$, or $-\infty < t < t_o$, involves m variables $x_1, x_2, \ldots, x_m$ and time t, or in all m+1 independent variables. The variables may be discrete or continuous. We may have two time series in m dimensions and their interdependence illustrated as follows:

$$z_1(t) = f_1(x_1,x_2,\ldots,x_m;t)$$
$$z_2(t) = f_2(x_1,x_2,\ldots,x_m;t),$$

or their interdependence

$$z_1(t) = g_1(z_2,x_1,x_2,\ldots,x_m;t)$$
$$z_2(t) = g_2(z_1,x_1,x_2,\ldots,x_m;t).$$

In general there may be n time series in m dimensions, $n<m$, namely $z(t)=f(x;t)$, $z(t)$ and $f(x;t)$ vectors with n components, and x a vector with m components. Similarly there may be any case of dependence among the z_i. We do not invoke stationarity here, although it is a helpful condition to assume at the onset. It should be clear then that these definitions include all the usual cases of time series in zero dimensions.

Our definition may be compared with the remarks of Hannan (1970) pages 94-95: "So far we have considered vector random processes x(t) in which t varies over the real line or some subset of it. As mentioned in Chapter I, in some applications t would in fact be a space variable, for example distance downstream from some fixed point on a river where x(t) might have three components corresponding to the "velocity" of the river. ...We shall use v again to indicate the point in the plane so that v needs two coordinates to name it. Again time variation may be present so that x(v,t) could be considered. Now the argument of x(.) varies over three dimensional space." According to our definition, this is a time series in two dimensions, or three independent variables. We could also have the situation of m dimensions, and p time variables conceptually: $\{x_i; t_j\}$ i=1,2,...,m, j=1,2,...,p.

From another point of view we may have x variables which may act as time variables when a time variable t_i does not occur. However t is a special type of variable.

Our definitions and models of time series in m dimensions have little or no overlap with the treatment of Bartlett (1975). Bartlett (1975, p.vii) states: "We may divide problems of spatial pattern (in contrast with complete random chaos) into (i) detecting departures from randomness, (ii) analysing such departures when detected, for example, in relation to some stochastic model and (iii) special problems which require separate consideration; for example, sophisticated problems of pattern recognition in specific fields, such as the computer reading of handwriting or recognition of chromosomes." Our main purpose is the generalization of the prediction models of Box and Jenkins from zero to m dimensions.

3. Two Physical Time Series in M Dimensions

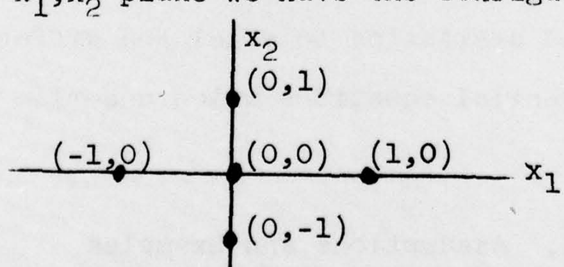
I shall give two time series in m dimensions: one, the characteristics of the sun, and the second, characteristics of a river. Almost all texts on time series consider the number of sunspots over time, z(t)=f(t) a zero dimensional time series. A time series in three dimensions, m=3, relates the number of sun spots to regions on the sun, each region given by spherical coordinates on the sun say (ξ_1, ξ_2, ξ_3). Thus $z(t)$, the number of sunspots at time t and position (ξ_1, ξ_2, ξ_3), where $z(\xi_i, t)$ may either be autoregressive, moving average, or a combination of the two.

The river characteristics may be described as a time series either in 1, or 2 or 3 dimensions. If x_{1i} represents the pollution at the center of the river at point x_{1i} over the whole length of the river, then the pollution of the river is given by $z(x_{1i}, t)$ in this case. If x_{2i} represents pollution at the width of the river at x_{1i} and x_{3i} the depth of the river, we obtain $z(x_{1i}, x_{2i}, x_{3i}, t)$, a time series in three dimensions. We do not necessarily assume stationarity in z in either case. One of the series characterize the number of sun spots on the sun and the other the pollution of a river in much clearer fashion than is possible by any time series in zero dimensions. These models show obviously that one may use time series and statistics to model the differential and partial differential equations which underlie these physical phenomena.

4. Assumptions and Examples

We assume essentially the same conditions on z as Box and Jenkins. F(z) the distribution of z, f(z) the density function, $\mu_z=0$, the mean, σ_z^2, the variance, $\sigma_{ij:z}$, the covariance function for $z(t_i)$ and $z(t_j)$, where t_j-t_i is a constant, all exist. We further assume z is stationary. Later Box and Jenkins drop the assumptions of stationarity which we shall also do. We need to state our assumptions introduced by the dependence of z on the variables $\{x_1,x_2,x_3,\ldots,x_m\}$. For simplicity we consider the case of m=2, with the variables $\{x_1,x_2\}$. In this framework we measure z at (x_1,x_2) or $z(x_1,x_2)$. The distribution of z, F(z) already defined applies over the x_1,x_2-plane. This distribution is also the same over the set formed by $\{x_1,x_2,t\}$, $-\infty<x_1<\infty$, $-\infty<x_2<\infty$, $-\infty<t<\infty$, although in practice we may only know z for $t\le t_o$. What is of greatest interest is the covariance structure in the x_1,x_2 plane as compared with the covariance structure as time changes. This will be made clearer in our discussion of one simple model, the moving average, MA.

In the x_1,x_2 plane we have the configuration



(0,0), (1,0), (-1,0), (0,1), and (0,-1).

The simple moving average model MA we shall use is:

$$z_{xyt}+\theta_{001}a_{xyt-1}+\theta_{101}a_{x-1yt-1}-\theta_{-101}a_{x+1yt-1} \qquad (4.1)$$

$$+\theta_{011}a_{xy-1t-1}+\theta_{0-11}a_{xy+1t-1}=a_{xyt}$$

We replace x_1 by x and x_2 by y for notational convenience. If we let all θ's except θ_{001} equal zero we get the zero dimensional Markov result

$$z_t + \theta_1 a_{t-1} = a_0 \qquad (4.2)$$

a model which ignores the contributions from the other points in the x_1, x_2 plane. The a's are independently distributed with mean zero and variance σ_a^2. Thus

$$\sigma_z^2 = \sigma_a^2(1+\theta_{001}^2+\theta_{101}^2+\theta_{-101}^2+\theta_{011}^2+\theta_{0-11}^2). \qquad (4.3)$$

Other conditions are found involving the coefficients of correlation in the x_1, x_2 plane and the correlations in time. The correlations are between the z's in the x_1, x_2 plane at time t-1; and the correlation between z_{xyt} and z_{xyt-1}. These are ρ_{001}, $\rho_{01\cdot}$, $\rho_{10\cdot}$, $\rho_{20\cdot}$, $\rho_{02\cdot}$; all the other six lag correlations are either equal to these or determined by these. The other relations needed for the determination of the five θ's are:

$$\sigma_z^2\rho_{001} = -\sigma_a^2\theta_{001} \qquad (4.4.1)$$

$$\sigma_z^2\rho_{10\cdot} = \sigma_a^2\theta_{001}(\theta_{101}+\theta_{-101}) \qquad (4.4.2)$$

$$\sigma_z^2\rho_{01\cdot} = \sigma_a^2\theta_{001}(\theta_{011}+\theta_{0-11}) \qquad (4.4.3)$$

$$\sigma_z^2\rho_{20\cdot} = \sigma_a^2\,\theta_{101}\theta_{-101} \qquad (4.4.4)$$

$$\sigma_z^2\rho_{02\cdot} = \sigma_a^2\theta_{0-11}\theta_{011}. \qquad (4.4.5)$$

These equations with (4.3) determine the θ's:

$$\theta_{001} = -\sigma_z^2/\rho_{001}\sigma_a^2 \tag{4.5.1}$$

$$\theta_{101} = \{k_1+(k_1^2-4k_2)^{\frac{1}{2}}\}/2 \tag{4.5.2}$$

$$k_1 = -\rho_{10.}/\rho_{001}, \quad k_2 = \sigma_z^2\rho_{20.}/\sigma_a^2$$

$$\theta_{-101} = \{k_1 - (k_1^2-4k_2)^{\frac{1}{2}}\}/2$$

$$\theta_{011} = \{k_3+(k_3^2-4k_4)^{\frac{1}{2}}\}/2 \tag{4.5.3}$$

$$k_3 = -\rho_{01.}/\rho_{001}, k_4 = \sigma_z^2\rho_{02.}/\sigma_a^2 \tag{4.5.4}$$

$$\theta_{0-11} = \{k_3 - (k_3^2-4k_4)^{\frac{1}{2}}\}/2 \tag{4.5.5}$$

The other six autocorrelations may be obtained as functions of these. The conditions $k_1^2 - 4k_2 \geq 0$, $k_3^2-4k_4 \geq 0$, do not always hold in the real domain. Only if we assume the errors a_{10-1} and a_{-101} equal; and a_{011} and a_{0-11} equal may we avoid these restrictions. If $k_1^2 -4k_2=0$, $\theta_{101} = \theta_{-101}$, and if $k_3^2-4k_4=0$, $\theta_{011}=\theta_{0-11}$.

What conditions must the coefficients of correlation statisfy? In this simple example $-1 \leq \rho_{001} \leq 1$, $-1 \leq \rho_{01.} \leq 1$, $-1 \leq \rho_{10.} \leq 1$, $\rho_{02.} \leq 1-2\rho^2_{01.}$ and $\rho_{20.} \leq 1-2\rho^2_{10.}$. Certain special cases may be found by eliminating the points (0,1),(0,-1) if θ_{011} and θ_{0-11} are both zeros, but the problem must be reanalyzed if the point (0,0) is missing, and θ_{001} is zero. Other correlations are given by:

$$\sigma_z^2\,\rho_{110} = \sigma_a^2(\theta_{-101}\theta_{0-11}+\theta_{011}\theta_{101}) \tag{4.6.1}$$

$$\sigma_z^2\,\rho_{-110} = \sigma_a^2(\theta_{101}\theta_{0-11}+\theta_{-101}\theta_{011}). \tag{4.6.2}$$

With equations (4.4) the autocorrelation function is completed. The eleven non-zero covariances are:

$$E(z_{00.}z_{0-1.}) = E(z_{00.}z_{01.}) = \sigma_z^2\rho_{01.},$$

$$E(z_{01.}z_{0-1.}) = \sigma_z^2\rho_{02.},\ E(z_{-10.}z_{10.}) = \sigma_z^2\rho_{20.},$$

$$E(z_{-10.}z_{00.}) = E(z_{00.}z_{10.}) = \sigma_z^2\rho_{10.}$$

$$E(z_{01.}z_{10.}) = E(z_{-10.}z_{0-1.}) = \sigma_z^2\rho_{110},$$

$$E(z_{-10.}z_{01.}) = E(z_{0-1.}z_{10.}) = \sigma_z^2\rho_{-110}, \text{ and}$$

$$E(z_{00-1}z_{000}) = \sigma_z^2\rho_{001}\ .$$

The invertibility conditions are found by inversion of (4.1). The restrictions on the constants θ, and the expression of (4.1) as an infinite autoregressive time series in x_1, x_2 and t are given in the paper of Aroian, Voss, Oprian . Discussion of autoregressive series, AR, are given in Aroian, **and Taneja,** and interrelationships between AR and AM, and the combined ARMA are given in Aroian, Oprian, Voss and Taneja.

5. Problems and Prospects

Are there problems in m dimensional time series for m>0, which may or mav not involve a single time parameter? In such cases relationships in all variables are considered without respect to the time variable. Consider the variable z the percent of a mineral available at discrete positions x_{1i}, x_{2i}, x_{3i}. Here there is no time parameter and z may be considered stationary as i varies between $-\infty$ and ∞. The stepsize i would vary for each dimension. A fundamental problem in geology then is to evaluate z at (x_{1i}, x_{2i}, x_{3i}) and to delineate the extent of a mine. In meteorology we have the temperature z, dependent on x_{1i}, x_{2i}, t_i, where x_i, y_i are the coordinates of place and t_i the time. In place of the temperature we may have barometric pressure, amount of rain, and other variables. We may think of a river face where observations are taken at fixed points at different times. Geometrically we view the x_1, x_2 plane moving along the time axis t, a two dimensional time series. We may be in a three-dimensional time series if we have additionally the height or depth of an observation. Or we may have two time series z_1 and z_2, two storms which are interacting at (x_1, x_2, x_3). This is a case of two time series in three dimensions. Let us look briefly at a problem in reading scores. We may consider reading scores z of children at time t dependent on coordinates x_{1i}, y_{2i}; two variables influencing reading which may be considered discrete and ordered in some way. We may consider z to be the height of a child as influenced by x_1, x_2, x_3 and t variables at discrete points, x_{1i}, x_{2i}, and x_{3i} and their relationship to time - and looking for cycles or for increase or decrease in the main variable z_i as influenced by the x_i. These models include important ones in biological sciences in environmental problems, air pollution, problems in evolution and in medicine since time is an important variable in all of these subjects. Thus we see a very wide need and envisage an equally wide use for this statistical technique. It should bring more understanding to these fields which presently are being considered in a general way without the benefits of mathematical and statistical insights. In agriculture z may be the yield of a crop at place (x_1, x_2) influenced by variables x_3, x_4 and x_5, a typical problem which is still not well handled without the time variable t. In fact some geographers have already been doing this particularly in space without the time variable considered. Time series will be more informative than the usual geographic block charts in different colors or designs. Other examples are earthquakes, hurricanes, and storms in geology and meteorology.

There is also the domain of history, the arts, sociology, anthropology and their relationships to time t. All of the preceding discussion may be summarized in the single equation

$$z_j = f(x_{1i}, x_{2i}, \ldots, x_{(m-1)i}, x_{mi}, t), -\infty < t < \infty, \text{ or } -\infty < t < t_o.$$

The appropriate joint probability distribution is assumed for the particular variate values of the variables. Researches in this field are clearly of the greatest importance and urgency.

6. Correlation Analysis Versus Time Series

We may consider time series in m dimensions as simply problems in correlation theory with the appropriate variables and variate values in a particular case. However if our interest is in cyclical situations, stationary time series or time series which may be transformed into stationary time series by use of appropriate operators, are much to be preferred. They provide useful models. The two methods should give the same final results but the time series approach should give more insight dependent on the particular model being chosen - autoregressive, moving average or a mixture of these two. Our results will be linear in the variable - which may eliminate interesting results in the nonlinear case. However, nonlinear time series either stationary or not should also be considered. In other papers we have investigated the autoregressive model in m dimensions, the moving average, a mixture of these, and the resulting problems of definition, estimation, and the relationships of these models to each other. This work will involve the autocorrelation function in m dimensions, the cross correlation function, and the multiple coefficient of correlation function, all in m dimensions.

7. Conclusion

Time series in m-dimensions have been defined. Some examples show the importance of the subject. The interrelationships among n time series in m dimensions are very briefly discussed for $n \geq 1$ and $m \geq 1$. Particular problems to be investigated are m dimensional moving average and autoregressive time series and their mixtures. One AM model is discussed at length as an example. The paper has purposely been general in order to obtain a very broad view of this subject and its scientific importance. The relationship between correlation theory and time series is briefly examined. The wide applicability of n time series in m dimensions series to the social sciences, the biological sciences, history, anthropology, economics, environmental problems, and meteorology is indicated.

ACKNOWLEDGMENTS

We appreciate the partial support of the Office of Naval Research, under contract ONR N00014-77-C-0438 and the Faculty Research Fund of Union College and University. We appreciate the comments of Professors Peter Bloomfield of Princeton University and Larry Haugh of the University of Vermont. They brought the papers of Bennett (1975), and the review paper of Cliff and Ord (1975) to our attention. Bennett (1975) has generalized the Box-Jenkins time series to spatial analysis, m = 2; his methods generalize the m dimensions under proper restrictions. He also generalizes the results of Akaike (1973). Bennetts's results are complementary to ours and apply to autoregressive models whether stationary or nonstationary.

BIBLIOGRAPHY

Akaike, H. (1973). Markovian representation of stochastic processes by canonical variables. SIAM J. of Control, 11.

Anderson, T.W. (1971). The Statistical Analysis of Time Series. New York: John Wiley & Sons, Inc.

Bennett, R.J. (1975). The representation and identification of spatio-temporal systems; an example of population diffusion in North-West England. Institute of British Geographers, Transactions, 66, 73-94.

Box, G.E.P. & Jenkins, G.M. (1976). Time Series Analysis: Forecasting and Control, revised edition. San Francisco: Holden-Day, Inc.

Campbell, D., Schaeffer, D.J., Aroian, L.A. (1977). Pollution in the Chicago Diversion Channel--An Example.

Cliff, A.D., & Ord, J.K. (1975). Model building and the analysis of spatial pattern in human geography. J. Royal Statist. Soc., Series B, 37, 297-328.

Davis, J.C., and McCullagh, M.J., editors (1975). Display and Analysis of Spatial Data. New York: John Wiley & Sons, Inc.

Hannan, E.J. (1970). Multiple Time Series. New York: John Wiley & Sons, Inc.

Jenkins, J.M., & Watts, D.J. (1968). Spectral Analysis and its Applications. San Francisco: Holden-Day, Inc.

Makridakis, S.A. (1956). Survey of time series. International Statist. Rev., 44, 1, 29-70.

Oprian, C., Taneja, V., Voss, D., & Aroian, L.A. (1977). General considerations and interrelationships between MA and AR models, time series in m-dimensions, the ARMA model.

Voss, D.A., Oprian, C.A., & Aroian, L.A. (1977). Moving average models--time series in m-dimensions.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 14 AES-7801

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle): 6 Time Series in M-Dimensions: Definition, Problems and Prospects.

5. TYPE OF REPORT & PERIOD COVERED: 9 Technical Report

6. PERFORMING ORG. REPORT NUMBER: AES-7801

7. AUTHOR(s): 10 Leo A. Aroian

8. CONTRACT OR GRANT NUMBER(s): 15 N00014-77-C-0438

9. PERFORMING ORGANIZATION NAME AND ADDRESS: Institute of Administration & Management, Union College and University, Schenectady, New York 12308

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 11 15 Jan 78

11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research Statistics & Probability Program, Office of Naval Research, Arlington, VA 22217

12. REPORT DATE: Jan. 15, 1978

13. NUMBER OF PAGES: 16 12 20p.

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unclassified

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Presented at a special session at the American Statistical Association annual meeting, August 1977, at Chicago, Ill.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

time series m-dimensions, moving average models, time series applications, time series examples, mine an example, river an example, storm an example, examples time series

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We define time series in m dimensions x, as follows: the observed variable z depends on $x_{1i}, x_{2i}, \ldots, x_{mi}$ and $t_i \leq t$ or $z = f(x_{1i}, x_{2i}, \ldots, x_{mi}, t_i)$, and similarly n time series in m variables $z_1, z_2, \ldots, z_n$, $z = f(x)$ where $z = f(x,t)$ and x are vectors This is the discrete case. The continuous case is similar.
(over)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

408 725

JOB

Distinction is made between m time series in zero dimension, all on the line, and one time series in m dimensions.